

①⑨ RÉPUBLIQUE FRANÇAISE  
INSTITUT NATIONAL  
DE LA PROPRIÉTÉ INDUSTRIELLE  
PARIS

①⑪ N° de publication : **2 764 611**

(à n'utiliser que pour les  
commandes de reproduction)

②① N° d'enregistrement national : **97 07225**

⑤① Int Cl<sup>6</sup> : C 12 N 15/12, C 12 N 5/10, 15/85, C 12 Q 1/68, C 07 K  
14/47, 16/18

⑫

## DEMANDE DE BREVET D'INVENTION

**A1**

②② Date de dépôt : 11.06.97.

③⑦ Priorité :

④③ Date de mise à la disposition du public de la  
demande : 18.12.98 Bulletin 98/51.

⑤⑥ Liste des documents cités dans le rapport de  
recherche préliminaire : *Se reporter à la fin du  
présent fascicule*

⑥⑦ Références à d'autres documents nationaux  
apparentés :

⑦① Demandeur(s) : FONDATION JEAN DAUSSET CEPH  
— FR.

⑦② Inventeur(s) : NERI CHRISTIAN et CANN HOWARD  
M.

⑦③ Titulaire(s) :

⑦④ Mandataire(s) : REGIMBEAU.

⑤④ SEQUENCES D'ADN RICHES EN TRIplet REPETE UTILES DANS LE DIAGNOSTIC DE MALADIES A  
REPETITION TRINUCLEOTIDIQUE.

⑤⑦ La présente invention concerne une séquence d'ADN  
transcrit et riche en triplet répété CAG/ CTG ou CGG/ GCC  
correspondant aux séquences A à L selon le tableau ainsi  
que leurs allèles et les séquences complémentaires.

Ces séquences sont utiles notamment dans le diagnos-  
tic des maladies à répétition trinucleotidique.

FR 2 764 611 - A1



La présente invention concerne la mise en évidence de gènes humains comportant des séquences à triplets répétés CAG/CTG ou CGG/GCC, ainsi que l'utilisation de ces gènes dans le diagnostic et l'éventuel traitement de certaines maladies monogéniques ou maladies polygéniques à composante héréditaire.

5 Le polymorphisme anormal de séquences à triplets répétés CAG/CTG ou CGG/GCC constitue une mutation impliquée dans au moins dix maladies humaines héréditaires (appelées maladies à triplet répété) qui sont principalement, mais non exclusivement, des maladies neurodégénératives. L'expansion (plusieurs triplets surnuméraires chez les malades) de séquences à triplet répété CAG/CTG est  
10 impliquée notamment dans l'atrophie spinobulbaire (SBMA), la dystrophie myotonique (DM), l'ataxie spinocérébelleuse (SCA) SCA1, SCA2, SCA3, et SCA6, l'atrophie dendato-rubro-pallidoluysienne (DRPLA), et la maladie de Huntington (1 à 3). L'expansion de séquences à triplet répété CGG/GCC est notamment impliquée dans le syndrome du X fragile (1). Enfin, le polymorphisme  
15 restreint de séquences à triplet répété CAG/CTG (par exemple un seul triplet surnuméraire chez les malades) est notamment impliquée dans la maladie de Behçet, une maladie autoimmune pour laquelle le passage de 5 à 6 codons GCT codant pour une polyalanine est associé à la maladie (4).

L'expansion des CAG/CTG ou CGG/GCC (appelée également mutation  
20 dynamique) est souvent associée à un phénomène d'anticipation, la taille des répétitions étant souvent corrélée de façon inverse avec l'âge de la survenue et/ou la sévérité des symptômes.

L'expansion des répétitions CAG/CTG peut apparaître dans les régions non codantes (DM) ou codantes (par exemple SBMA, HD, DRPLA, SCAs) des  
25 transcrits. Ces transcrits sont parfois exprimés dans les tissus autres que le cerveau et lorsqu'ils sont traduits conduisent à des produits de gènes plus grands et portant un domaine polyglutamine anormalement allongé (1 à 3).

L'expansion de triplets répétés CAG/CTG peut être très importante (par exemple DM ; plusieurs centaines de triplets), modérée (par  
30 exemple HD ou SCA1 ; en général plus de 35 triplets), voire très modérée (par exemple SCA6 ; en général moins de 35

triplets). A l'état normal, ces répétitions peuvent être hautement polymorphiques (par exemple HD ou SCA1) ou faiblement à très faiblement polymorphiques (par exemple SCA2 et SCA6) (1 à 3). Une variation d'un seul triplet à l'intérieur d'un triplet répété court et faiblement polymorphique peut aussi causer une maladie (4).

- 5 La mise en évidence d'expansion CAG/CTG dans l'ADN génomique de patients (5) et/ou l'anticipation dans les familles à risque a suggéré que les mutations dynamiques sont impliquées dans SCA4, SCA5, l'ataxie cérébrale dominante autosomale (ADCA de type 2, aussi désignée sous le sigle SCA7), et dans les formes familiales de la psychose maniaco-dépressive (PMD) et de la schizophrénie (6, 7). Dans le cas de SCA7, la mise en évidence d'expansions de polyglutamine a suggéré que le triplet répété impliqué est du type triplet répété CAG dans une région codante (8).

- 15 L'autisme et la démence familiale qui montrent des variabilités dans l'âge de la survenue ou dans la sévérité des symptômes pourraient également être causées par des mutations dynamiques.

Un assez grand nombre de maladies pourraient également avoir pour origine ou impliquer des mutations dynamiques :

- la maladie de Parkinson,
- les paraplégies spastiques,
- 20 - les ataxies cérébrospinales non répertoriées précédemment,
- les cataractes zonulaires,
- les tremblements incontrôlables,
- les neuropathies amyloïdes familiales,
- les arthrites granulomateuses familiales (maladie de Blau),
- 25 - les microsomies hémifaciales,
- certaines anémies et glaucomes,
- les désordres obsessionnels.

Ce qui précède démontre l'importance considérable des triplets répétés pour le diagnostic et le traitement de plusieurs maladies.

La présente invention repose sur une étude systématique des répétitions CAG/CTG ou CGG/GCC (qui seront ci-après désignées par "[CAG]*n*" et "[CGG]*n*" respectivement, *n* étant le nombre de répétitions du triplet), cette étude ayant été réalisée avec des banques d'ADNc humains, celles-ci ayant subi un premier criblage  
5 général, puis les séquences retenues ayant été sélectionnées sur la base d'un certain nombre de critères permettant de s'assurer que les séquences en cause appartenaient à de nouveaux gènes humains et pouvaient, avec une grande probabilité, être impliquées dans des maladies à triplet répété. Enfin, les séquences sélectionnées ont fait l'objet d'une étude plus complète destinée à permettre leur  
10 localisation.

Dans le cadre de la présente invention, deux ensembles d'ADNc provenant de cerveaux humains ont été étudiés pour analyser la présence de répétitions CAG en utilisant l'hybridation d'oligonucléotides sur des membranes à haute densité. Les deux librairies ont été obtenues à partir de ARNms à l'aide d'amorce oligo-dT,  
15 l'une des librairies étant constituée de clones d'ADNc de cerveau foetal (FB) et l'autre librairie étant constituée de clones d'ADNc normalisés de cerveau de nouveau-né (NIB).

De façon générale, la présente invention repose sur la mise en évidence de certaines séquences susceptibles d'être impliquées dans les maladies à triplet répété  
20 obtenues dans des conditions d'hybridation permettant de sélectionner des ADNc qui contiennent un nombre de séquences [CAG] ou [CGG] supérieur à 7, lesquelles sont plus susceptibles d'être polymorphes dans une population malade.

Les conditions complètes d'analyse et de sélection de ces différentes séquences ne seront pas détaillées ci-après, seuls seront fournis les éléments  
25 permettant de les localiser et de les réisoler grâce, notamment, aux amorces PCR qui seront décrites ci-après.

Plus particulièrement, la présente invention concerne une séquence d'ADN transcrit riche en triplet répété CAG/CTG ou CGG/GCC correspondant aux séquences A à L du tableau, ainsi que leurs allèles normaux et mutés et les  
30 séquences complémentaires.

Par "allèles normaux" on entend désigner les allèles tels qu'ils ont été isolés ou tels qu'il peuvent être isolés de prélèvements d'individus normaux, les "allèles mutés" étant les allèles des gènes portant les séquences [CAG] $n$  ou [CGG] $n$  anormalement répétées.

- 5 Il est important de noter que, bien que certaines de ces séquences soient, totalement ou en partie, publiques, les séquences en cause sont pour la première fois identifiées comme faisant partie d'un gène pouvant présenter une mutation, lesdits gènes n'ayant, par ailleurs, en eux-mêmes jamais été décrits.

- Les séquences mises en évidence présentent des pourcentages  
10 d'hétérozygotie (HTZ) très variables allant de 0 à 0,75 %. Les séquences monomorphiques (pour lesquelles HTZ = 0) sont au moins autant susceptibles d'être impliquées dans des maladies à triplet répété que les séquences polymorphes.

- Bien entendu, comme cela est mentionné précédemment, ces séquences sont identifiées dans le tableau et pourront, si besoin est, être réisolées en utilisant  
15 les amorces suivantes : SEQ ID 1 à 24, les amorces 1 à 24 correspondant, par paire, aux séquences A à L mentionnées précédemment.

- Les séquences ainsi mises en évidence peuvent, tout d'abord, être utilisées dans le cadre du diagnostic, et plus exactement d'un pronostic. En effet, comme un grand nombre de maladies ayant un support génétique, la mise en évidence de la  
20 présence d'une séquence présentant un nombre de répétitions anormal de triplet [CAG] $n$  ou [CGG] $n$  ne peut en elle-même assurer de la survenue de la maladie, mais doit être interprétée en fonction d'un ensemble d'autres informations pour permettre, soit un diagnostic très précoce, soit éventuellement une surveillance spécifique, surtout dans les familles à risque.

- 25 Ce diagnostic peut être effectué en comparant la séquence d'ADN selon l'invention du patient avec une séquence normale pour détecter la présence de répétitions trinuéotidiques supplémentaires.

- Plus particulièrement, la présente invention concerne un procédé de mise en évidence des risques d'apparition d'une maladie à  
30 répétition trinuéotidique chez un patient, caractérisé en ce qu'on compare la séquence d'ADN conforme à l'invention dudit patient à

une séquence normale pour détecter la présence de répétitions trinuécléotidiques supplémentaires.

Bien entendu, il existe un grand nombre de méthodes qui permettent la mise en évidence de triplets surnuméraires par rapport à des séquences normales, par exemple il est possible de mettre en évidence des différences de poids moléculaires en utilisant des gels et une méthode de type RFLP. Mais, les méthodes les plus efficaces consistent à pratiquer préalablement à la mise en évidence des différences une opération d'amplification, par toute méthode appropriée, notamment la méthode dite "PCR", même si d'autres méthodes sont utilisables.

Il n'est pas nécessaire ici de décrire en détail la méthode PCR, celle-ci permet d'amplifier de façon considérable une séquence spécifique comprise entre deux séquences appelées "amorces".

Parmi les amorces utilisables dans le cadre des procédés selon l'invention, il faut citer les amorces des SEQ ID 1 à 24 qui constituent deux par deux des paires d'amorces pour chacune des séquences objet de l'invention.

En utilisant les amorces décrites précédemment, on amplifie les séquences selon la présente invention et il est alors possible, par comparaison avec un échantillon normal et/ou étalon, de mettre en évidence la présence des triplets surnuméraires et donc la possibilité de survenue d'une maladie liée à ce type de mutation dynamique.

Il est également intéressant de noter que certaines maladies à triplet répété sont connues pour être associées à des variations modérées à très faibles du nombre de triplets répétés, comme par exemple, SCA6 et la maladie de Behçet. Dans ces conditions, la méthode diagnostic peut permettre, non seulement un bon pronostic de la survenue de la maladie, mais également d'évaluer le moment où cette maladie surviendra et/ou son éventuelle sévérité.

La présente invention concerne également les gènes et leurs allèles qui portent, au moins en parties, ces séquences, lesdits gènes étant, bien entendu, impliqués directement dans la survenue de la maladie.

Elle a également pour objet une cellule transformée exprimant tout ou partie des susdits gènes, une protéine obtenue à partir d'une telle cellule, une protéine interagissant avec la protéine précédemment mentionnée ainsi qu'un vecteur exprimant l'une de ces protéines.

- 5 Les gènes correspondant pourront être exprimés dans des cellules par des moyens connus afin de produire les protéines correspondantes.

Il est possible d'envisager l'utilisation desdites protéines dans certains kits de diagnostic par exemple.

- 10 De façon générale, ces protéines étant impliquées dans la maladie, on souhaitera en diminuer la quantité, soit en bloquant leur expression par des méthodes appropriées au niveau des gènes ou des éléments de régulation, soit en les fixant ou en les inactivant, par exemple en utilisant des protéines réceptrices jouant le rôle de leurres. Là encore, ces protéines réceptrices ou inactivées pourront être générées in situ par expression des séquences d'ADN correspondantes.

- 15 Il est également possible de prévoir la réalisation d'anticorps monoclonaux correspondant à ces protéines afin d'envisager le blocage desdites protéines lorsque cela est souhaité, l'ensemble de ces opérations pouvant être réalisé directement in vivo par exemple, en utilisant des techniques de thérapie génique, en particulier en utilisant des vecteurs qui porteront les séquences d'expression des gènes.

- 20 On a mis en évidence le fait que les protéines présentant des domaines polyglutamines anormalement étendus avaient des propriétés d'aggrégation anormales, tant entre elles, qu'avec d'autres protéines (11, 12). Les aggrégats ainsi créés sont probablement impliqués dans la genèse des maladies à triplet répété, c'est pourquoi il est également possible de prévoir une thérapie dans laquelle les agents thérapeutiques viendraient empêcher l'aggrégation, soit en bloquant la
- 25 molécule comme décrit précédemment, soit en se fixant à la molécule pour empêcher le rapprochement avec d'autres protéines.

- On peut, dans le cadre de la thérapie, prévoir d'utiliser des variants complets ou délétés de ces protéines, normales ou non (quant au domaine [CAG]<sub>n</sub> ou [CGG]<sub>n</sub>).
- 30

Les séquences selon la présente invention peuvent être obtenues grâce aux informations figurant au tableau et aux références qui y sont données et en utilisant les séquences d'amorce qui sont décrites dans les identificateurs de séquence ci-joints.

5 Les deux librairies utilisées sont :

- une première librairie (5) contenant 60 000 ADNc non normalisés de cerveau foetal humain (clones FB) (Laboratoire du Dr. Hans Lehrach, Max Plank Institute for Molecular Genetics, Berlin, Allemagne) sous-clonés dans le vecteur p-SPORT-1 (Life Technologies, Inc., Gaithersburg, MA, USA), et
- 10 - la seconde librairie contient 40 128 ADNc normalisés de cerveau de nouveau-né (clones NIB) sous-clonés dans un vecteur lafmid (6) et qui est une partie des ressources du consortium IMAGE (Lawrence Livermore Lab., Livermore, CA,) et du programme EST
- 15 Merck WU (Washington University, St-Louis, LO, USA).

D'autre part, un certain nombre d'autres informations ont été recueillies en utilisant Genbank Database (NCBI Bethesda, MA, USA).

Les méthodes de sélection et d'analyse qui ont été mises en oeuvre ne font pas en elles-mêmes partie de la présente invention puisque l'invention a  
20 essentiellement pour objet les séquences ainsi obtenues et leur utilisation, notamment, dans des méthodes de diagnostic ou des méthodes de traitement thérapeutique.

Dans la présente analyse, on a retenu, non seulement les séquences [CAG] $n$  ou [CGG] $n$  parfaites, mais également les séquences [CAG] $n$  ou [CGG] $n$   
25 complexes, c'est-à-dire celles qui sont ponctuées par des insertions de triplets ; en effet cette présence de triplets insérés s'observe notamment dans le cas de SCA1, alors que tel n'est pas le cas pour SBMA, MD et HD.

D'autre part, les nouveaux clones sélectionnés à partir de librairies NIB sont distincts de ceux sélectionnés dans le groupe FB. Ceci est en accord avec le  
30 fait que le cerveau humain exprime un grand nombre de gènes distincts à des stades de développement différents.



Enfin, des présentes observations il ressort que des séquences [CAG]*n* ou [CGG]*n* sont rares dans les ADNc humains représentatifs des régions 3' des ARNm.

5 Bien que dans l'état actuel de la présente invention, les séquences qui constituent l'objet de l'invention n'aient pas été reliées directement à des affections héréditaires, la présence d'extensions anormales CAG ou CGG peut être reliée avec le risque de survenue d'une maladie, en particulier lorsqu'il y a une composante familiale.

10 Il existe un assez grand nombre de maladies qui actuellement n'ont pas été localisées et trouveront probablement à être reliées avec les séquences selon la présente invention.

Parmi les éléments additionnels qui ont été pris en compte pour étudier la pertinence des séquences retenues figure l'existence éventuelle d'un cadre de lecture ouvert (ORF) dans lequel les séquences [CAG]*n* ou [CGG]*n* codent pour  
15 une extension polyglutamine.

TABLEAU

Séquence	Nom du clone (CEPH)	Triplet 5'-3'	Hétérozygotie	Nombre d'allèles (nombre de triplets)	Localisation
A	1.45	(CTG) <sub>9</sub> *	aucune	1	14
B	2.52	(CAG) <sub>11</sub> *	aucune	1	18
C	2.99	(CAG) <sub>28</sub> *	aucune	1	X
D	2.122	(CTG) <sub>13</sub> *	aucune	1	6
E	3.58	(CTG) <sub>11</sub> *	aucune	1	7
F	i.10	(CAG) <sub>7</sub>	0.025	2 (7, 8)	1
G	i.12	(CTG) <sub>11</sub> *	0.75	2 (11, 13)	X
H	i.34	(CAG) <sub>9</sub> *	aucune	1	10, 16
I	i.111	(CTG) <sub>13</sub> *	aucune	1	X
J	i.129	(CTG) <sub>7</sub> (ATG) <sub>10</sub>	0.675	4 (17-20)	12
K	1.85	(CGG) <sub>14</sub> *	aucune	1	2
L	2.296	(GCC) <sub>14</sub> *	aucune	1	-

\* structure imparfaite

## LISTE DE SEQUENCES

## (1) INFORMATIONS GENERALES:

## (i) DEPOSANT:

- (A) NOM: FONDATION JEAN DAUSSET 6 CENTRE D'ETUDES DU  
POLYMORPHISME HUMAIN (CEPH)
- (B) RUE: 27 RUE JULIETTE DODU
- (C) VILLE: PARIS
- (E) PAYS: FRANCE
- (F) CODE POSTAL: 75010

(ii) TITRE DE L' INVENTION: SEQUENCES D'ADN RICHES EN TRIPLET REPETE  
UTILES DANS LE DIAGNOSTIC DE MALADIES A REPETITION  
TRINUCLEOTIDIQUE

(iii) NOMBRE DE SEQUENCES: 24

## (iv) FORME DECHIFFRABLE PAR ORDINATEUR:

- (A) TYPE DE SUPPORT: Floppy disk
- (B) ORDINATEUR: IBM PC compatible
- (C) SYSTEME D' EXPLOITATION: PC-DOS/MS-DOS
- (D) LOGICIEL: PatentIn Release #1.0, Version #1.30 (OEB)

## (2) INFORMATIONS POUR LA SEQ ID NO: 1:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 17 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

## (vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Max Plank Institute for Molecular Genetics,  
Laboratory of Dr hans Lehrach, Berlin, Allemagne
- (B) CLONE: ICRFp507K15223

## (ix) CARACTERISTIQUE:

- (A) NOM/CLE: 1.45

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 1:

CGCTCCTTCC TCTCAAG

17

## (2) INFORMATIONS POUR LA SEQ ID NO: 2:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 18 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

## (vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Max Plank Institute for Molecular Genetics,  
Laboratory of Dr hans Lehrach, Berlin, Allemagne
- (B) CLONE: ICRFp507K15223

- (ix) CARACTERISTIQUE:  
(A) NOM/CLE: 1.45

- (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 2:

CTCAGCATGG GGTAAG

18

- (2) INFORMATIONS POUR LA SEQ ID NO: 3:

- (i) CARACTERISTIQUES DE LA SEQUENCE:  
(A) LONGUEUR: 19 paires de bases  
(B) TYPE: nucléotide  
(C) NOMBRE DE BRINS: simple  
(D) CONFIGURATION: linéaire

- (ii) TYPE DE MOLECULE: ADNc

- (vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Max Plank Institute for Molecular Genetics,  
Laboratory of Dr hans Lehrach, Berlin, Allemagne  
(B) CLONE: ICRFp507M22190

- (ix) CARACTERISTIQUE:  
(A) NOM/CLE: 2.52

- (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 3:

CTACAACAGA AATTCCGAC

19

- (2) INFORMATIONS POUR LA SEQ ID NO: 4:

- (i) CARACTERISTIQUES DE LA SEQUENCE:  
(A) LONGUEUR: 16 paires de bases  
(B) TYPE: nucléotide  
(C) NOMBRE DE BRINS: simple  
(D) CONFIGURATION: linéaire

- (ii) TYPE DE MOLECULE: ADNc

- (vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Max Plank Institute for Molecular Genetics,  
Laboratory of Dr hans Lehrach, Berlin, Allemagne  
(B) CLONE: ICRFp507M22190

- (ix) CARACTERISTIQUE:  
(A) NOM/CLE: 2.52

- (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 4:

GGAACGACTG GAACAG

16

- (2) INFORMATIONS POUR LA SEQ ID NO: 5:

- (i) CARACTERISTIQUES DE LA SEQUENCE:  
(A) LONGUEUR: 21 paires de bases  
(B) TYPE: nucléotide  
(C) NOMBRE DE BRINS: simple  
(D) CONFIGURATION: linéaire

- (ii) TYPE DE MOLECULE: ADNc

## (vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Max Plank Institute for Molecular Genetics,  
Laboratory of Dr hans Lehrach, Berlin, Allemagne
- (B) CLONE: ICRF p507F12249

## (ix) CARACTERISTIQUE:

- (A) NOM/CLE: 2.99

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 5:

CATGGGGTTG CCGAAGAGGA G

21

## (2) INFORMATIONS POUR LA SEQ ID NO: 6:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 20 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADNc

## (vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Max Plank Institute for Molecular Genetics,  
Laboratory of Dr hans Lehrach, Berlin, Allemagne
- (B) CLONE: ICRFp507F12249

## (ix) CARACTERISTIQUE:

- (A) NOM/CLE: 2.99

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 6:

CGCCCTTCTG GAAGGCTCAG

20

## (2) INFORMATIONS POUR LA SEQ ID NO: 7:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 18 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADNc

## (vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Max Plank Institute for Molecular Genetics,  
Laboratory of Dr hans Lehrach, Berlin, Allemagne
- (B) CLONE: ICRFp507C15296

## (ix) CARACTERISTIQUE:

- (A) NOM/CLE: 2.122

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 7:

TGGCAGAATG GATTTGTG

18

## (2) INFORMATIONS POUR LA SEQ ID NO: 8:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 19 paires de bases

- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Max Plank Institute for Molecular Genetics,  
Laboratory of Dr hans Lehrach, Berlin, Allemagne
- (B) CLONE: ICRFp507C15296

(ix) CARACTERISTIQUE:

- (A) NOM/CLE: 2.122

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 8:

CCATCTTGAA GAAGTACTT

19

(2) INFORMATIONS POUR LA SEQ ID NO: 9:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 19 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Max Plank Institute for Molecular Genetics,  
Laboratory of Dr hans Lehrach, Berlin, Allemagne
- (B) CLONE: ICRFp507L13177

(ix) CARACTERISTIQUE:

- (A) NOM/CLE: 3.58

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 9:

TCACTCCTCC ATTGTCTGC

19

(2) INFORMATIONS POUR LA SEQ ID NO: 10:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 18 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Max Plank Institute for Molecular Genetics,  
Laboratory of Dr hans Lehrach, Berlin, Allemagne
- (B) CLONE: ICRFp507L13177

(ix) CARACTERISTIQUE:

- (A) NOM/CLE: 3.58

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 10:

CTGTCCGAGG AGGTATCG

18

(2) INFORMATIONS POUR LA SEQ ID NO: 11:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 18 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Reseau I.M.A.G.E., Lawrence Livermore National Laboratories, Livermore, CA, USA.
- (B) CLONE: 35570

(ix) CARACTERISTIQUE:

- (A) NOM/CLE: i.10

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 11:

ACCTTGTTCA GTCCATTG

18

(2) INFORMATIONS POUR LA SEQ ID NO: 12:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 21 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Reseau I.M.A.G.E., Lawrence Livermore National Laboratories, Livermore, CA, USA.
- (B) CLONE: 35570

(ix) CARACTERISTIQUE:

- (A) NOM/CLE: i.10

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 12:

GTTGTCTTTA TAAAACACAG A

21

(2) INFORMATIONS POUR LA SEQ ID NO: 13:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 17 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Reseau I.M.A.G.E., Lawrence Livermore National Laboratories, Livermore, CA, USA.
- (B) CLONE: 43014

- (ix) CARACTERISTIQUE:  
(A) NOM/CLE: i.12

- (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 13:

CCCAAGTCCC ACAATAG

17

- (2) INFORMATIONS POUR LA SEQ ID NO: 14:

- (i) CARACTERISTIQUES DE LA SEQUENCE:  
(A) LONGUEUR: 16 paires de bases  
(B) TYPE: nucléotide  
(C) NOMBRE DE BRINS: simple  
(D) CONFIGURATION: linéaire

- (ii) TYPE DE MOLECULE: ADNc

- (vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Reseau I.M.A.G.E., Lawrence Livermore National  
Laboratories, Livermore, CA, USA.  
(B) CLONE: 43014

- (ix) CARACTERISTIQUE:  
(A) NOM/CLE: i.12

- (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 14:

GCGTCAGTGG CAGAAG

16

- (2) INFORMATIONS POUR LA SEQ ID NO: 15:

- (i) CARACTERISTIQUES DE LA SEQUENCE:  
(A) LONGUEUR: 18 paires de bases  
(B) TYPE: nucléotide  
(C) NOMBRE DE BRINS: simple  
(D) CONFIGURATION: linéaire

- (ii) TYPE DE MOLECULE: ADNc

- (vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Reseau I.M.A.G.E., Lawrence Livermore National  
Laboratories, Livermore, CA, USA.  
(B) CLONE: 54662

- (ix) CARACTERISTIQUE:  
(A) NOM/CLE: i.34

- (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 15:

TTTTTATGAT CCTGGTTC

15

- (2) INFORMATIONS POUR LA SEQ ID NO: 16:

- (i) CARACTERISTIQUES DE LA SEQUENCE:  
(A) LONGUEUR: 18 paires de bases  
(B) TYPE: nucléotide  
(C) NOMBRE DE BRINS: simple  
(D) CONFIGURATION: linéaire

- (ii) TYPE DE MOLECULE: ADNc



## (vii) SOURCE IMMEDIATE:

(A) BIBLIOTHEQUE: Reseau I.M.A.G.E., Lawrence Livermore National Laboratories, Livermore, CA, USA.

(B) CLONE: 54662

## (ix) CARACTERISTIQUE:

(A) NOM/CLE: 1.34

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 16:

TGTGTTAGGT CCTAGAAG

18

## (2) INFORMATIONS POUR LA SEQ ID NO: 17:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 18 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADNc

## (vii) SOURCE IMMEDIATE:

(A) BIBLIOTHEQUE: Reseau I.M.A.G.E., Lawrence Livermore National Laboratories, Livermore, CA, USA.

(B) CLONE: 42315

## (ix) CARACTERISTIQUE:

(A) NOM/CLE: i.111

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 17:

CCTCCAGCAG GGTATGGC

18

## (2) INFORMATIONS POUR LA SEQ ID NO: 18:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 19 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADNc

## (vii) SOURCE IMMEDIATE:

(A) BIBLIOTHEQUE: Reseau I.M.A.G.E., Lawrence Livermore National Laboratories, Livermore, CA, USA.

(B) CLONE: 42315

## (ix) CARACTERISTIQUE:

(A) NOM/CLE: i.111

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 18:

CGAACGGACC TGGGAAC TG

19

## (2) INFORMATIONS POUR LA SEQ ID NO: 19:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 18 paires de bases

- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vii) SOURCE IMMEDIATE:

(A) BIBLIOTHEQUE: Réseau I.M.A.G.E., Lawrence Livermore National Laboratories, Livermore, CA, USA.

(B) CLONE: 52900

(ix) CARACTERISTIQUE:

(A) NOM/CLE: i.129

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 19:

TCTGCCCCTG GAGGTGGG

18

(2) INFORMATIONS POUR LA SEQ ID NO: 20:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 27 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vii) SOURCE IMMEDIATE:

(A) BIBLIOTHEQUE: Réseau I.M.A.G.E., Lawrence Livermore National Laboratories, Livermore, CA, USA.

(B) CLONE: 52900

(ix) CARACTERISTIQUE:

(A) NOM/CLE: i.129

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 20:

CTTCACTGGC GCTTTTCTT CAGCTTC

27

(2) INFORMATIONS POUR LA SEQ ID NO: 21:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 19 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vii) SOURCE IMMEDIATE:

(A) BIBLIOTHEQUE: Max Plank Institute for Molecular Genetics, Laboratory of Dr Hans Lehrach, Berlin, Allemagne

(B) CLONE: ICRFp507P04189

(ix) CARACTERISTIQUE:

(A) NOM/CLE: 1.85

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 21:

GAACAGCCTA TCTCATTCC

19

(2) INFORMATIONS POUR LA SEQ ID NO: 22:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
  - (A) LONGUEUR: 32 paires de bases
  - (B) TYPE: nucléotide
  - (C) NOMBRE DE BRINS: simple
  - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Max Plank Institute for Molecular Genetics,  
Laboratory of Dr hans Lehrach, Berlin, Allemagne
- (B) CLONE: ICRFp507P041188

(ix) CARACTERISTIQUE:

- (A) NOM/CLE: 1.85

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 22:

TCAGAAGTAA CTTGAATAAA TAATCATATT CG

32

(2) INFORMATIONS POUR LA SEQ ID NO: 23:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
  - (A) LONGUEUR: 16 paires de bases
  - (B) TYPE: nucléotide
  - (C) NOMBRE DE BRINS: simple
  - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Max Plank Institute for Molecular Genetics,  
Laboratory of Dr hans Lehrach, Berlin, Allemagne
- (B) CLONE: ICRFp507B16262

(ix) CARACTERISTIQUE:

- (A) NOM/CLE: 2.296

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 23:

CACTTGGAAT CCACAC

16

(2) INFORMATIONS POUR LA SEQ ID NO: 24:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
  - (A) LONGUEUR: 16 paires de bases
  - (B) TYPE: nucléotide
  - (C) NOMBRE DE BRINS: simple
  - (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(vii) SOURCE IMMEDIATE:

- (A) BIBLIOTHEQUE: Max Plank Institute for Molecular Genetics,  
Laboratory of Dr hans Lehrach, Berlin, Allemagne
- (B) CLONE: ICRFp507B16262

(ix) CARACTERISTIQUE:  
(A) NOM/CLE: 2.296

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 24:

AGAGATCCTG AAGGAC

16

REFERENCES

1. Paulson, H.L. and Fishbeck, K.H. Trinucleotide repeats in neurogenetic disorders *Annu Rev NeuroSci* 19, 79-107 (1996).
- 5 2. Zoghbi, H. The expanding world of ataxins. *Nature Genet.* 14, 237-238 (1996).
3. Zuchenko O, Bailey J, Bonnen P, Ashizawa T, Stockton DW, Amos C, Dobyns WB, Subramony SH, Zoghbi HY, and Lee CC. Autosomal dominant cerebellar ataxia (SCA6) associated with small glutamine expansion in the alpha 1A-voltage-dependent calcium channel. *Nature Genet* 15, 62-69 (1997).
- 10 4. Mizuki, N. et al. Triplet repeat polymorphism in the transmembrane region of the MICA gene: a strong association of six GCT repetitions with Behçet disease. *Proc Natl Acad Sci USA* 94, 1298-1303 (1997).
- 15 5. Schalling M, Hudson TJ, Buetow KH, Housman DE. Direct detection of novel expanded trinucleotide repeats in the human genome. *Nature Genet* 4, 135-139, 1993.
6. O'Donovan MC, Guy C, Craddock N, Murphy KC, Cardno AG, Jones LA, Owen MJ McGuffin P. Expanded CAG repeats in schizophrenia and bipolar disorders. *Nature Genet* 10, 380-381 (1995).
- 20 7. Morris AG, Gaitonde E, McKenna PJ, Mollon JD, Hunt DM. CAG repeat expansions and schizophrenia : association with disease in females and with early age-at-onset. *Hum Mol Genet* 4, 1957-1961 (1995).
8. Trottier, Y. et al. Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxia. *Nature* 378, 403-406 (1995).
- 25 9. Meier-Ewert, S., Maier, E., Ahmadi, A., Curtis, J. and Lebrach, H. An Automated approach to generating expressed sequence catalogues. *Nature* 361, 375 (1993).

10. Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L. & Efstratiadis, A. Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci. USA* 91, 9228-9232 (1994).
11. Haad, D., Perry, M.J. and Haynes, L. Cellular transglutaminases in neural development. *Int. J. Devl Neuroscience* 11, 709-720 (1993).
12. Li X-J, Li S-H, Sharp AH, Nucifora Jr FC, Schilling G, Lanahan A, Worley P, Snyder SH, Ross CA. A Huntingtin-associated protein enriched in brain with implications for pathology. *Nature* 378, 398-402 (1995).

REVENDICATIONS

- 1) Séquence d'ADN transcrit riche en triplet répété CAG/CTG ou CGG/GCC correspondant aux séquences A à L selon le tableau ainsi que leurs  
5 allèles normaux et mutés et les séquences complémentaires.
- 2) Séquence d'ADN, caractérisée en ce qu'elle porte au moins en partie une séquence selon la revendication 1.
- 3) Procédé de mise en évidence des risques d'apparition d'une maladie à répétition trinuécléotidique chez un patient, caractérisé en ce qu'on  
10 compare la séquence d'ADN selon la revendication 1 dudit patient à une séquence normale pour détecter la présence de répétitions trinuécléotidiques supplémentaires.
- 4) Procédé selon la revendication 3, caractérisé en ce qu'on effectue une amplification de tout ou partie des séquences selon la revendication 1 et que l'on met en évidence dans le produit d'amplification la présence de répétitions  
15 trinuécléotidiques supplémentaires.
- 5) Procédé selon la revendication 4, caractérisé en ce que l'amplification des séquences est effectuée avec les amorces décrites SEQ ID 1 à 24 correspondant aux séquences A à L respectivement.
- 6) Cellule transformée exprimant tout ou partie d'un gène selon la  
20 revendication 2.
- 7) Protéine obtenue à partir d'une cellule transformée selon la revendication 6.
- 8) Protéine interagissant avec une protéine selon la revendication 7.
- 9) Anticorps monoclonal correspondant à une protéine selon l'une  
25 des revendications 7 ou 8.
- 10) Vecteur exprimant une protéine selon l'une des revendications 7 ou 8.
- 11) Utilisation d'une protéine selon l'une des revendications 7 ou 8, d'un anticorps selon la revendication 9 ou d'un vecteur selon la revendication 10  
30 pour la préparation d'un médicament.

INSTITUT NATIONAL

de la

PR PRIETE INDUSTRIELLE

## RAPPORT DE RECHERCHE

## PRELIMINAIRE

établi sur la base d e dernières r vendications  
déposées avant le commencement de la rech rcheN° d'enregistrement  
nationalFA 546202  
FR 9707225

DOCUMENTS CONSIDERES COMME PERTINENTS		Revendications concernées de la demande examinée
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	
X	WO 94 28172 A (MEDICAL RES COUNCIL ;KNIGHT SAMANTHA JAYNE LESLEY (GB); DAVIES KAY) le document en entier, esp. page 15,ligne 19- page 18 ligne 13 ---	1-11
X	WO 95 01437 A (UNIV MINNESOTA) le document en entier, esp. revendications et exemples ---	1-11
X	WO 95 25179 A (UNIV MASSACHUSETTS MEDICAL) revendications ---	1-3
X	WO 93 17104 A (MASSACHUSETTS INST TECHNOLOGY) methods et revendications ---	1-4
X	YING-HUI FU ET AL: "VARIATION OF THE CGG REPEAT AT THE FRAGILE X SITE RESULTS IN GENETIC INSTABILITY: RESOLUTION OF THE SHERMAN PARADOX" CELL, vol. 67, no. 6, 20 décembre 1991, pages 1047-1058, XP000371773 * le document en entier * ---	1-4
X	WO 92 20825 A (BAYLOR COLLEGE MEDICINE ;UNIV EMORY MED (US)) revendications ---	1-4
X	PUJANA M. A. ET AL., : "Cloning (CAG/GTC)n STSs by an ALu-(CAG/GTC)n PCR method: an approach to human chromosome 12 and spinocerebellar ataxia 2 (SCA2)" NUCLEIC ACIDS RESEARCH, vol. 24, no. 18, - 1996 pages 3651-3652, XP002059375 * le document en entier * -----	1-4
		DOMAINES TECHNIQUES RECHERCHES (Int.CL.6)
		C12Q C07K
Date d'achèvement de la recherche		Examineur
18 mars 1998		Müller, F
CATEGORIE DES DOCUMENTS CITES		
T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant		
X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : pertinent à l'ensemble d'au moins une revendication ou arrière-plan technologique général O : divulgation non-caricature P : document intercalaire		

EPO FORM 1503 (2.02) (P04C13)

1